DOCUMENT RESUME

ED 366 914 CS 011 591

AUTHOR McEneaney, John E.
TITLE Does Parallel Distributed Processing Provide a Plausible Framework for Modeling Early Reading Acquisition?
PUB DATE 4 Apr 94
NOTE 7p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Beginning Reading; *Cognitive Processes; Comparative Analysis; Elementary Education; *Mathematical Models; Reading Research; Reading Skills
IDENTIFIERS Emergent Literacy; *Parallel Distributed Processing

ABSTRACT

A study compared a parallel distributed processing (PDP) model with a more traditional symbolic information processing model that accounts for early reading acquisition by human subjects. Two experimental paradigms were simulated. In one paradigm (a "savings" paradigm) subjects were divided into two groups and trained with two sets of stimuli: consistent orthographic representation of voicing, or inconsistent mapping of orthography onto voicing. A total of 32 simulated trials were generated with each model. Analysis of the savings paradigm data for the symbolic model revealed no significant main or interaction effects. Analysis of the PDP data revealed significant main and interaction effects. A second paradigm involved a forced-choice task testing subjects' ability to make use of analytic generalization. A total of eight simulated trials were generated with each model. Analysis indicated no significant difference between the symbolic model and chance performance, while the PDP model had perfect performance, indicating that the network had learned to make use of the orthography-voicing relationship implicit in the stimuli. Findings suggest that, in the domain of early reading acquisition, the problem with the PDP model approach is not that it is too weak but that it is too strong--even the simplest PDP models exhibit learning beyond what is observed in human subjects faced with similar learning tasks. (Contains 18 references. An appendix of data is attached.) (RS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

A Paper Presented at the Annual Meeting of the American Educational Research Association
April 4, 1994 - Session 1.34, Reading in Children

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Does parallel distributed processing provide a plausible framework for modeling early reading acquisition?

John E. McEneaney
Indiana University South Bend
JMcEnean@Indiana.edu

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. McEneaney
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED 366 914

Parallel Distributed Processing (PDP) models are central to a great deal of current research in the cognition of reading and are beginning to assert an influence well beyond the boundaries of the cognitive science research community. Adams (1990), for example, has grounded an argument for a specific approach to beginning reading instruction on a PDP model of word learning (Seidenberg & McClelland, 1989). Ehri (1992) has begun to adopt PDP concepts in explaining the development of sight word reading, and researchers with interests in learning difficulties (Seidenberg, 1992; Patterson, Seidenberg, & McClelland, 1989) have begun to model disabilities through simulated "lesioning" of working PDP systems.

There is evidence, however, that the currently dominant PDP models of reading (McClelland & Rumelhart, 1981, 1988 pp. 203-239; Seidenberg & McClelland, 1989) are missing essential aspects of the cognition of early word perception and learning. One consistent general problem has been that the distributed representation of knowledge in PDP models is difficult to reconcile with the acquisition and use of distinct perceptual units (e.g. onsets and rimes) that appear to play an important role in learning to read (Treiman, 1992; Goswami, 1986, 1988). Another problem concerns the data sets used in testing the adequacy of PDP models which have relied almost exclusively on mature adult readers even in models that make explicitly developmental claims (e.g. The Seidenberg & McClelland (1989) developmental model cites 29 empirical studies only 1 of which employed children as subjects.) But neither of these limitations are theoretically critical in the sense that they rule out the PDP approach as an appropriate framework for models of early literacy acquisition.

Recent research in reading acquisition suggests, however, that a critical test of the adequacy of PDP models in explaining early reading acquisition may be at hand. As a consequence of a series of carefully controlled studies Byrne (1992) has identified what he calls the "default acquisition procedure" for reading. Briefly, these studies (Byrne, 1984, 1992; Byrne & Carroll, 1989) investigated the way children learn to read new orthographies and found that, in the absence of explicit training in orthographic analysis, children adopt a non-analytic (i.e. paired associate) approach. Despite being provided with a completely regular sound-symbol system, the children did not induce phoneme-grapheme correspondences over extended periods of training. Moreover, in a series of related studies exploring the learning of new orthographies by adult subjects the same default procedure was evident at a sub-phonemic level. Although adult subjects recognized the alphabetic character of the orthography at the syllable and word levels, they did not discover the underlying system of regularity between the orthography and sub-phonemic feature elements including the voicing contrast which has been shown to have high perceptual salience (Miller & Nicely, 1955).

CS 011 091

The present investigation reports on two simulations of the Byrne studies that suggest the PDP framework is incapable of accounting for the default acquisition procedure. The studies reported here compare a PDP processing model similar to the Seidenberg & McClelland (1989) model (developed using the McClelland & Rumelhart (1988) BP simulation) with a more traditional symbolic information processing model (McEneaney, 1991, 1992). Results of trials simulating the Byrne studies indicate that the symbolic model provides a more adequate treatment of early reading acquisition. Theoretical analysis suggests that the results of the simulations are not specific to either model parameters or learning rules employed in the simulations but that these differences constitute a genuinely critical test between a symbolic and a PDP approach in accounting for early reading acquisition by human subjects.

Two experimental paradigms were simulated in the present investigation. In one paradigm (labelled by Byrne & Carroll a "savings" paradigm) subjects were divided into two groups and trained with two sets of stimuli. One group of subjects was trained using stimulus sets within which voicing was consistently represented orthographically. The second group of subjects was trained with stimulus sets that employed an inconsistent mapping of orthography onto voicing. If subjects were using analytic learning it was reasoned that subjects in the consistent group would have an advantage over subjects in the inconsistent group. A second paradigm involved a forced-choice task testing subjects' ability to make use of analytic generalization. Subjects who had learned letters in a new orthography were asked to identify novel stimuli that retained one critical orthographic marker (indicating voicing) from the original training set. If analytic learning had occurred, the one critical feature would have been enough to allow subjects to make the correct choice since the distractor in the forced-choice task did not include the correct voicing feature. As noted above, both adults and children learning the new orthographies failed to exhibit analytic learning in either of the experimental tasks. The same was <u>not</u> true across the simulations carried out in the present investigation.

The savings paradigm study began by training both models to a level of accuracy that ensured 100% correct recognition over at least three passes through the stimulus set used in the first episode of training. Since two different stimulus sets were employed, the set used in the first episode was counter balanced. When the models had achieved the performance criterion, a second episode of training was initiated that employed either a second consistent or a second inconsistent stimulus set. A total of 32 simulated trials were generated with each model. Data analysis involved a 2 (Episode) X 2 (Consistency) repeated measures design that employed trials to criterion as the dependent measure, a design identical to that reported by Byrne & Carroll (1989, p. 313).

Analysis of the savings paradigm data for the symbolic model revealed no significant main ($F_{Episode} = 0.3659$, $p = 0.55$; $F_{Consistency} = 1.2$, $p = 0.28$) or interaction effects ($F_{Interaction} = 0.3659$, $p = 0.55$). Analysis of the PDP data on the other hand revealed significant main ($F_{Episode} = 22.615$, $p < .0001$; $F_{Consistency} = 75.0721$, $p < .0001$)and interaction effects ($F_{Interaction} = 60.1093$, $p < .0001$) all of which were significant at $p < .0001$. Inspection of means for the PDP data from the second episode of training revealed both an advantage for the consistent group and a disruption of learning for the inconsistent group in the second episode of training, indicating that the PDP model had learned analytically.

The forced choice simulations also began by training both models to a level of accuracy that ensured 100% correct recognition over three passes through a training set with a regular mapping of orthography onto 2 sub-phonemic features (place of articulation and voicing). When the learning criterion was achieved the models were presented with 8 new stimuli made up of a new orthographic feature along with the previously seen feature indicating either voiced or voiceless pronunciation. They were also presented with two new phonemes one voiced and the other unvoiced (e.g. /f/ and /v/) and required to select the phoneme represented by the orthographic stimulus. If analytic learning was occurring it was reasoned that performance on this forced-choice task would be significantly greater than chance (i.e. > 4/8 correct).

A total of 8 simulated trials were generated with each model. Data analysis involved a t-test to determine whether a significant deviation from chance performance had occurred (as in experiment 1, Byrne, 1984). Analysis indicated no significant difference between the symbolic model and chance performance ($T = 0.8864$, p (1-tailed) $= .1972$). The PDP model, however, had perfect performance (8/8 correct) across all 8 simulated trials indicating that the network had learned to make use of the orthography-voicing relationship implicit in the stimuli.

Although the analysis indicates that the PDP model learns analytically it might be argued that this performance is an artifact of implementation specific parameters rather than a more general characteristic of the PDP framework. In the present case, however, it turns out that two characteristics of the training and testing data sets are more important than the parameters of the model. One is that these data sets conform to the linear predictability constraint (any one feature can be predicted by a linear combination of the activation of the other features) which means these data sets are guaranteed to be learnable (McClelland & Rumelhart, 1986). The second characteristic is that the one-to-one mapping of orthographic and phonemic feature elements results in a an auto-associative learning task that is inevitably analytic. Even the simplest kinds of PDP networks (e.g. a 2-layer perceptron or a 1-layer auto-associative net) demonstrate the same analytic learning following training. Altering parameters will change the rate at which learning occurs (or can prevent learning), but <u>if learning does occur, it will be analytic</u>.

In the domain of early reading acquisition, therefore, the problem with the PDP approach is not that it is too weak (Rumelhart & McClelland, 1986) but that it is too <u>strong</u>. Even the simplest PDP models exhibit learning beyond what we observe in human subjects faced with similar learning tasks. This, of course, does not eliminate PDP models from consideration at other stages of development but it does raise two interesting questions about the role of PDP models in the cognition of reading. If PDP systems become one of a set of components, what do the other components look like and how do the components interact (if at all)? Whatever the final verdict regarding the place of PDP models in explaining reading acquisition one thing seems clear: the results reported in this paper rule out PDP processing as an explanatory framework for reading acquisition in its earliest stages.

References

Adams, M. (1990). Beginning to read. Cambridge, MA: MIT Press.

Byrne, B. (1984). On teaching articulatory phonetics via an orthography. Memory & Cognition, 12(2), 181-189.

Byrne, B. (1992). Studies in the acquisition procedure for reading: Rationale, hypotheses, and data. In Gough, Ehri, & Treiman (Eds), Reading acquisition. Hillsdale, NJ: Lawrence Erlbaum.

Byrne, B. & Carroll, M. (1989). Learning artificial orthographies: Further evidence of a non-analytic acquisition procedure. Memory & Cognition, 17(3), 311-317.

Ehri, L. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. Gough, L. Ehri & R. Treiman (Eds), Reading acquisition. Hillsdale, NJ: Lawrence Erlbaum.

Goswami, U. (1986). Children's use of analogy in learning to read: A developmental study. Journal of Experimental Child Psychology, 42, 73-83.

Goswami, U. (1988). Orthographic analogies and reading development. Quarterly Journal of Experimental Psychology, 40A, 239-268.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1 An account of basic findings. Psychological Review, 88, 375-407.

McClelland, J. L. & Rumelhart, D. E. (1986). A distributed model of human learning and memory. In James McClelland & David Rumelhart (Eds.), Parallel Distributed Processing, Volume 2: Psychological and biological models. Cambridge, MA: MIT Press.

McCLelland, J. L., & Rumelhart, D. E. (1988). Explorations in parallel distributed processing. Cambridge, MA: MIT Press.

McEneaney, J. (1991). A unification-based model of perceptual learning in reading: Part 1. Logical foundations. Journal of Artificial Intelligence in Education, 2(4), 51-69.

McEneaney, J. (1992). A unification-based model of perceptual learning in reading: Part 2. Computational Implementation. Journal of Artificial Intelligence in Education, 3(1), 81-104.

Miller, G. & Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America, 27, 338-352.

Patterson, K., Seidenberg, M., & McClelland, J. (1989). Connections and disconnections:

acquired dyslexia in a computational model of reading processes. In R. Morris (Ed.), Parallel distributed processing. Implications for psychology and neurobiology. Oxford: Clarendon Press.

Rumelhart, D. E. & McCLelland, J. L. (1986). PDP models and general issues in cognitive science. In David Rumelhart & James McClelland (Eds.), Parallel Distributed Processing. Volume 1: Foundations. Cambridge, MA: MIT Press.

Seidenberg, M. & McClelland, J. (1989). A distributed, developmental model of word recognition and naming. Psychological Review, 96(4), 523-568.

Seidenberg, M. (1992). Dyslexia in a computational model of word recognition in reading. In P. Gough, L. Ehri, & R. Treiman (Eds), Reading acquisition. Hillsdale, NJ: Lawrence Erlbaum.

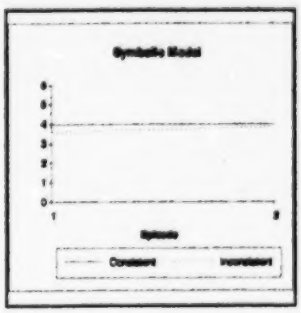
Treiman, R. (1992). The role of intrasyllabic units in learning to read and spell. In P. Gough, L. Ehri, & R. Treiman (Eds), Reading acquisition. Hillsdale, NJ: Lawrence Erlbaum.

Savings Paradigm

Factor	Levels	
Episode	2	Episode factor is within subjects
Consistency	2	Consistency factor is across subjects

Symbolic Model

Source	DF	F	p
Consistency	1	1.2000	0.2820
Episode	1	0.36585	0.5498
Interaction	1	0.36585	0.5498

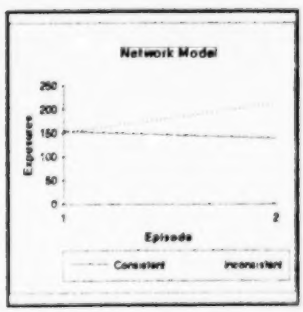


Means and (standard deviations):

	Episode	
Consistency	Trial 1	Trial 2
Consistent	4.000	4.000
	(0.7303)	(0.9666)
Inconsistent	3.625	3.875
	(0.8062)	(0.9574)

Network Model

Source	DF	F	p
Consistency	1	75.072	<0.0001
Episode	1	22.6105	<0.0001
Interaction	1	60.1093	<0.0001



Means and (standard deviations):

Consistency	**Episode**	
	Trial1	Trial2
Consistent	153.312	138.500
	(5.48597)	(11.7473)
Inconsistent	131.000	212.812
	(11.1535)	(32.2701)

Forced-choice Paradigm

Population T Test	**Mean Correct**	**Variance**	**T**	**p (1-tailed)**
Population	4.00	1.69697		
Symbolic Model	4.33	1.69697	0.886405	0.1972
Network Model	8.00	0	NA	NA